JOURNAL OF MULTIVARIATE ANALYSIS **57**, 277–296 (1996)
ARTICLE NO. 0034

Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes

EFSTATHIOS PAPARODITIS

University of Cyprus, Nicosia, Cyprus

We consider an r-dimensional multivariate time series $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ which is generated by an infinite order vector autoregressive process. We show that a

tribution of the growing set of estimated autoregressive coefficients and to the corresponding set of estimated moving average coefficients (impuls responses).

1. INTRODUCTION

Bootstrapping time series data has received considerable attention in recent years. In contrast to the case of independent data, several different approaches for bootstrapping stationary observations are available. In particular, in case the parametric structure of the process considered is assumed to be known, the natural approach is to incorporate that structure into the algorithm used to generate the bootstrap replicates. A well-studied case here is that of a finite parameter autoregressive moving average model; cf. Bose [3], Kreiss and Franke [6]. However, when the process considered is generated by a model with unknown parametric structure, then model-free techniques are required in order to resample the dependence of data and to evaluate the distribution of the statistics of interest. For this case alternatives like the blockwise bootstrap Künsch [7] and Liu and

Received September 27, 1994; revised September 1995.

AMS 1990 subject classifications: primary 62M10, 62E20, 62G09; secondary 62P20.

Key words and phrases: bootstrap, parameter estimates, infinite order vector autoregressions, autoregressive coefficients, moving average coefficients.

0047-259X/96 $18.00

Singh [9] as well as generalizations thereof, Politis and Romano [12] have been proposed.

A different situation occurs, however, when the precise form of the parametric model generating the data is not assumed to be known, but it is assumed that this model belongs to a certain, much more general class of parametric models. A practically relevant case here is the one in which it is assumed that the true model belongs to the class of stationary linear processes having an autoregressive representation of infinite order. In this context and in the univariate case, Kreiss [5] in his habilitation thesis investigated properties of a bootstrap procedure which works by generating replicates of the observed data using an estimated finite parameter autoregressive process, the order of which increases at some appropriate rate with the sample size. A similar procedure has been proposed by Paparoditis and Streitberg [11] in order to estimate the distribution of certain generalized second-order statistics which are useful in model identification.

In this paper we consider an application of the above bootstrap procedure for approximating the joint distribution of the (with sample size growing) set of estimated autoregressive coefficients as well as that of the corresponding moving average coefficients, in cases where the stochastic process generating the data has an infinite order vector autoregressive representation. Our discussion begins by developing multivariate generalizations of some of the basic univariate results of Kreiss [5] and apply these to the problem of estimating the joint distribution of the parameter estimators of interest. We show that if the auto-regressive order fitted to the data increases at some appropriate rate with the sample size, the bootstrap estimators obtained are at least asymptotically as sound as the conventional large sample Gaussian approximations; i.e., we establish the so-called asymptotic validity of the bootstrap proposal. It is worth noting here that interest in the distribution of the parameter estimators in the moving average representation of the system arises frequently in applied work, e.g., in econometrics, where these coefficients are interpreted as the impulse responses of the system; cf. Lütkepohl [10].

The paper proceeds as follows. Section 2 briefly outlines the resampling procedure proposed and discusses for the multivariate case considered here, some basic results on bootstrapping infinite dimensional vector autoregressive processes via finite order autoregressive fitting. Section 3 discusses the properties of the procedure applied to the distribution of parameter estimators and establishes its asymptotic validity in approximating the distribution of the estimators in the autoregressive and in the moving average representations of the process. Section 4 contains proofs of the theorems obtained.

2. Resampling the Linear Structure by Vector Autoregressive Fitting

Consider a second-order, weak stationary, nondeterministic r-dimensional stochastic process $\{\mathbf{y}_t, t \in \mathbf{Z}\}$, with an autoregressive representation given by

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{A}_j \mathbf{y}_{t-j} + \varepsilon_t, \tag{1}$$

where the $\mathbf{A}_j = (a_{lc,j})_{l,c=1,2,\ldots,r}$ are real $r \times r$ matrices. In this representation $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{rt})^\top$ is an r-dimensional sequence of independent identically distributed (i.i.d.) random vectors with distribution function denoted by F, where $E(\boldsymbol{\varepsilon}_t) = 0$, $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s^\top) = \delta_{ts} \boldsymbol{\Sigma}$ for all $t, s \in \mathbf{Z}$ and $E\,|\varepsilon_{it}\,\varepsilon_{jt}\,\varepsilon_{lt}\,\varepsilon_{st}| = \gamma_4 < \infty$. The $r \times r$ matrix $\boldsymbol{\Sigma}$ is positive definite. We assume that $\det(\mathbf{A}(z)) \neq 0$ for all $|z| \leqslant 1$, where $\mathbf{A}(z) = \mathbf{I}_r - \sum_{j=1}^{\infty} \mathbf{A}_j z^j$ and $\mathbf{I}_r$ denotes the $r \times r$ identity matrix. In the following we consider the class of processes defined as in (1) and for which the parameter matrices satisfy the condition that a constant $\eta > 0$ exists such that $\sum_{j=1}^{\infty} \|\mathbf{A}_j\| (1+\eta)^j < \infty$, where $\|\cdot\|$ denotes Schur's matrix norm defined by $\|\mathbf{A}_j\|^2 = \mathrm{tr}(\mathbf{A}_j^\top \mathbf{A}_j)$. The subordinate matrix norm associated with the Euclidean norm and which is defined by $\|\mathbf{A}\|_1 = \sup_{\mathbf{x} \neq 0} \{\|\mathbf{A}\mathbf{x}\| / \|\mathbf{x}\|\}$ is also sometimes used in the following. The class of processes considered includes the finite parameter vector autoregressive moving average processes as a special case.

It is well known that under these conditions $\{\mathbf{y}_t, \mathbf{Z}\}$ may be also expressed as a causal infinite order moving average process, i.e., $\mathbf{y}_t = \varepsilon_t + \sum_{j=1}^{\infty} \mathbf{B}_j \varepsilon_{t-j}$, where we have $\mathbf{B}(z) = \mathbf{A}(z)^{-1} = \mathbf{I}_r + \sum_{j=1}^{\infty} \mathbf{B}_j z^j$, $\sum_{j=1}^{\infty} \|\mathbf{B}_j\| < \infty$, and $\mathbf{B}_j = (b_{lc,})_{l,c=1,2,\ldots,r}$, $j = 1, 2, \ldots,$ are the corresponding $r \times r$ coefficient matrices. These are given by $\mathbf{B}_j = \mathbf{J}_k \boldsymbol{\Pi}_k^j \mathbf{J}_k^\top$, where $\mathbf{J}_k$ denotes the $r \times kr$ matrix $\mathbf{J}_k = (\mathbf{I}_r \mathbf{0} \cdots \mathbf{0})$, $\mathbf{0}$ is the $r \times r$ null matrix, and the $kr \times kr$ block matrix $\boldsymbol{\Pi}_k$ is defined by

$$\boldsymbol{\Pi}_k = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \cdot & \mathbf{A}_k \\ \mathbf{I}_r & \mathbf{0} & \cdots & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r & \cdots & \cdot & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_r & \mathbf{0} \end{pmatrix}. \tag{2}$$

In applications the elements of the matrix $\mathbf{B}_j$ are interpreted as the impulse response coefficients of the system.

Suppose that a time series $\mathbf{y}_t$, $t=1, 2, ..., T$, has been observed and that the covariance structure of the process has been parametrized by a finite parameter, k-order vector autoregressive process. The order k of the fitted autoregression is a function of the sample size; i.e., $k=k(T)$, where k increases, at some rate to be discussed later, simultaneously with T. Let $\mathbf{A}(k)=(\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_k)$ be the matrix of the first k autoregressive parameter matrices in the representation (1) and denote by $\hat{\mathbf{A}}(k)=(\hat{\mathbf{A}}_{1,k}, \hat{\mathbf{A}}_{2,k}, ..., \hat{\mathbf{A}}_{k,k})$ the corresponding Yule–Walker estimator given by

$$(\hat{\mathbf{A}}_{1,k}, \hat{\mathbf{A}}_{2,k}, ..., \hat{\mathbf{A}}_{k,k})=\hat{\mathbf{\Gamma}}_{1,k}^{\top}\hat{\mathbf{\Gamma}}_{k}^{-1}, \tag{3}$$

where $\hat{\mathbf{\Gamma}}_{1,k}=(T-k)^{-1}\sum_{t=k}^{T-1}\mathbf{Y}_{t,k}\mathbf{y}_{t+1}^{\top}$, $\hat{\mathbf{\Gamma}}_{k}=(T-k)^{-1}\sum_{t=k}^{T-1}\mathbf{Y}_{t,k}\mathbf{Y}_{t,k}^{\top}$, and $\mathbf{Y}_{t,k}=(\mathbf{y}_t^{\top}, \mathbf{y}_{t-1}^{\top}, ..., \mathbf{y}_{t-k+1}^{\top})^{\top}$. Denote by $\tilde{\boldsymbol{\varepsilon}}_{k,t}$ the residuals of this model; i.e., $\tilde{\boldsymbol{\varepsilon}}_{k,t}=\mathbf{y}_t-\sum_{j=1}^{k}\hat{\mathbf{A}}_{j,k}\mathbf{y}_{t-j}$, where $t=k+1, k+2, ..., T$ and let $\hat{\mathbf{\Sigma}}_k$ be the estimator of $\mathbf{\Sigma}$ calculated by $\hat{\mathbf{\Sigma}}_k=(T-k)^{-1}\sum_{t=k+1}^{T}\hat{\boldsymbol{\varepsilon}}_{k,t}\hat{\boldsymbol{\varepsilon}}_{k,t}^{\top}$, with $\hat{\boldsymbol{\varepsilon}}_{k,t}=\tilde{\boldsymbol{\varepsilon}}_{k,t}-\bar{\boldsymbol{\varepsilon}}_k$ and $\bar{\boldsymbol{\varepsilon}}_k=(T-k)^{-1}\sum_{t=k+1}^{T}\tilde{\boldsymbol{\varepsilon}}_{k,t}$. Furthermore, denote by $\hat{\mathbf{B}}(k)=(\hat{\mathbf{B}}_{1,k}, \hat{\mathbf{B}}_{2,k}, ..., \hat{\mathbf{B}}_{k,k})$ the corresponding estimator of $\mathbf{B}(k)=(\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_k)$ given by $\hat{\mathbf{B}}_{j,k}=\mathbf{J}_k\hat{\mathbf{\Pi}}_k^j\mathbf{J}_k^{\top}$. The matrix $\hat{\mathbf{\Pi}}_k$ appearing in this expression is the one obtained after replacing in the matrix $\mathbf{\Pi}_k$ the unknown matrices $\mathbf{A}_j$ by their estimators $\hat{\mathbf{A}}_{j,k}$.

Now, let $\{\mathbf{l}(k)\}_{k\in\mathbf{N}}$ be an arbitrary sequence of $kr^2\times 1$ vectors satisfying the condition $0<M_1\leqslant\|\mathbf{l}(k)\|^2\leqslant M_2<\infty$, and let $\mathbf{a}(k)=\mathrm{vec}(\mathbf{A}(k))$, $\hat{\mathbf{a}}(k)=\mathrm{vec}(\hat{\mathbf{A}}(k))$, $\mathbf{b}(k)=\mathrm{vec}(\mathbf{B}(k))$, and $\hat{\mathbf{b}}(k)=\mathrm{vec}(\hat{\mathbf{B}}(k))$, where $\mathrm{vec}(\cdot)$ denotes the vector stacking operator. For k and $T\to\infty$ asymptotic normality of the statistic $(T-k)^{1/2}\mathbf{l}(k)^{\top}(\hat{\mathbf{a}}(k)-\mathbf{a}(k))$ has been established by Lewis and Reinsel [8] under the assumption $k^3/T\to 0$ as $k, T\to\infty$. Note that the condition $\sqrt{T}\sum_{j=k+1}^{\infty}\|\mathbf{A}_j\|\to 0$ needed by these authors is satisfied here because $\|\mathbf{A}_j\|\to 0$ at a geometric rate as $j\to\infty$. Under the same set of conditions, asymptotic normality of the statistic $T^{1/2}(\hat{\mathbf{b}}(m)-\mathbf{b}(m))$ for $1\leqslant m\leqslant k$ fixed and where $\mathbf{b}(m)=\mathrm{vec}(\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_m)$ and $\hat{\mathbf{b}}(m)=(\hat{\mathbf{B}}_{1,k}, \hat{\mathbf{B}}_{2,k}, ..., \hat{\mathbf{B}}_{m,k})$ has been also established; cf. Lütkepohl [10].

In this paper the distribution of the statistics $(T-k)^{1/2}\,\mathbf{l}(k)^{\top}\,(\hat{\mathbf{a}}(k)-\mathbf{a}(k))$ and $T^{1/2}\mathbf{l}(k)^{\top}(\hat{\mathbf{b}}(k)-\mathbf{b}(k))$ is estimated using the following bootstrap procedure:

1. Depending on the assumptions imposed on F, and for $s\in\mathbf{N}$ given, an i.i.d. sample $\boldsymbol{\varepsilon}_t^*$, $t=-s+1, -s+2, ..., 1, 2, ..., T$, is first generated. For instance, if F is assumed to belong to a finite dimensional parametric family of distributions, i.e., $F\in\{F(;, \boldsymbol{\theta}): \boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbf{R}^m\}$, then the $\boldsymbol{\varepsilon}_t^*$ can be obtained as i.i.d. sequences from $F(;, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ stays for a consistent estimator of $\boldsymbol{\theta}$. Alternatively, and in the absence of any a priori assumptions on the distribution of the true errors, the series $\boldsymbol{\varepsilon}_t^*$ can be obtained as an i.i.d.

sample from the empirical distribution function $\hat{F}_T(\cdot)$ which puts mass $(T-k)^{-1}$ on each of the centered residuals vectors $\hat{\boldsymbol{\varepsilon}}_t$.

2. Using this noise series, bootstrap replicates $\{\mathbf{y}_t^*, t=1, 2, ..., T\}$ are obtained by

$$\mathbf{y}_t^* = \sum_{j=0}^{t+s-1} \hat{\mathbf{B}}_{j,k} \boldsymbol{\varepsilon}_{t-j}^*, \tag{4}$$

where the matrices $\hat{\mathbf{B}}_{j,k}$, $j=1, 2, ...,$ are easily calculated using the formula

$$\hat{\mathbf{B}}_{j,k} = \sum_{i=1}^{j} \hat{\mathbf{B}}_{j-i,k} \hat{\mathbf{A}}_{i,k}, \tag{5}$$

with $\hat{\mathbf{A}}_{j,k}=0$ for $j>k$ and $\hat{\mathbf{B}}_{0,k}=\mathbf{I}_r$. Denote then by $\hat{\mathbf{A}}^*(k)=(\hat{\mathbf{A}}_{1,k}^*, \hat{\mathbf{A}}_{2,k}^*, ..., \hat{\mathbf{A}}_{k,k}^*)$ and $\hat{\mathbf{B}}^*(k)=(\hat{\mathbf{B}}_{1,k}^*, \hat{\mathbf{B}}_{2,k}^*, ..., \hat{\mathbf{B}}_{k,k}^*)$ the estimates of the autoregressive and of the moving average parameters, respectively, which are obtained by fitting VAR(k) process to the series $\{\mathbf{y}_t^*, t=1, 2, ..., T\}$.

3. Steps 1 and 2 are repeated a large number of times and the distribution of the statistics considered is approximated by the empirical distribution of the bootstrap quantities $(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))$ and $T^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{b}}^*(k)-\hat{\mathbf{b}}(k))$, where $\hat{\mathbf{a}}^*(k)=\mathrm{vec}(\hat{\mathbf{A}}^*(k))$ and $\hat{\mathbf{b}}^*(k)=\mathrm{vec}(\hat{\mathbf{B}}^*(k))$.

Before looking in more detail at the properties of the above procedure, some basic results which are frequently used in the following are first discussed.

The following theorem is a multivariate generalization of a univariate result given by Berk [1]. It follows using Eq. (2.8) of Lewis and Reinsel [8, p. 397] and the property $\|\mathbf{A}_j\| \leqslant C\rho^j$, where $C>0$ and $\rho \in (0, 1)$ are constants.

THEOREM 2.1. *Let* $\{\hat{\mathbf{A}}(k)\}_{k\in\mathbf{N}}$ *be the sequence of estimators of* $\mathbf{A}(k)$, *where* $\hat{\mathbf{A}}(k)$ *is given in* (3). *If* $k\to\infty$ *as* $T\to\infty$ *such that* $k/T^{1/2}\to 0$ *then*

$$\|\hat{\mathbf{A}}(k)-\mathbf{A}(k)\| = O_P\left(\frac{k^{1/2}}{T^{1/2}}\right).$$

The next two theorems are multivariate analogues of univariate results given by Kreiss [5]. Their proofs which follow essentially the same lines as the proofs of their univariate counterparts are briefly outlined in Section 4.

THEOREM 2.2. *Let* $k/T^{1/2}\to 0$ *and denote by* $\{\hat{\mathbf{B}}_{j,k} j\in\mathbf{N}\}_{k\in\mathbf{N}}$ *the sequence of the coefficient matrices in the infinite series expansion* $(\mathbf{I}_r - \sum_{j=1}^{k} \hat{\mathbf{A}}_{j,k} z^j)^{-1}$, *if this expansion exists. Then a constant $C \in (0, \infty)$ exists such that for all $j \in \mathbf{N}$,*

$$\|\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j\| \leqslant C(1+k^{-1})^{-j} \frac{k}{T^{1/2}}.$$

Consistency of $\hat{\mathbf{A}}(k)$ as an estimator of $\mathbf{A}(k)$ has been established in Theorem 1 of Lewis and Reinsel [8] under the assumption that $k^2/T \to 0$ as $k, T \to \infty$. Under the same assumption it is easily seen that by Theorem 2.2 above, $\|\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j\| \to 0$ in probability. Furthermore, using the definition of $\|\cdot\|$ we have that $\|\hat{\mathbf{B}}(k) - \mathbf{B}(k)\| \to 0$ in probability if $k^3/T \to 0$ as k and $T \to \infty$; i.e., $\hat{\mathbf{B}}(k)$ is a consistent estimator for the sample size increasing parameter set $\mathbf{B}(k)$.

THEOREM 2.3. *Let $\mathbf{w}_1$ and $\mathbf{w}_2$ be two r-dimensional random vectors with $\mathbf{w}_1 \sim \hat{F}_T$ and $\mathbf{w}_2 \sim F_T$, where F_T denotes the empirical distribution function of the true errors ε_t, $t = k+1, k+2, ..., T$. If $k/T^{1/2} \to 0$ as k and $T \to \infty$, then*

$$E\|\mathbf{w}_1 - \mathbf{w}_2\|^2 = O_P\left(\frac{k^2}{T}\right).$$

Using the above theorem the asymptotic properties of the empirical distribution function $\hat{F}_T$ of the centered residuals $\hat{\boldsymbol{\varepsilon}}_{k,t}$ as an estimator of the distribution function F of the true errors are easily established. For this, the distance between any two distributions on the space $\mathscr{M} = \{P : P$ probability measure on $\mathscr{B}$, $\int \|\mathbf{x}\|^2 \, dP < \infty\}$ is measured in the following by Mallow's metric which for $P_1, P_2 \in \mathscr{M}$ is defined by $d_2(P_1, P_2) = \inf\{E\|\mathbf{x}_1 - \mathbf{x}_2\|^2\}^{1/2}$. The infimum is taken here over all real valued random vectors $(\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$ with $\mathscr{L}(\mathbf{x}_1) = P_1$ and $\mathscr{L}(\mathbf{x}_2) = P_2$ and where $\mathscr{L}(\mathbf{x})$ denotes the law of $\mathbf{x}$; cf. Bickel and Freedman [2] for details. Now, because $d_2(\hat{F}_T, F) \leqslant d_2(\hat{F}_T, F_T) + d_2(F_T, F)$ we get using Lemma 8.4 of Bickel and Freedman [2] and Theorem 2.3 the following result.

THEOREM 2.4. *If $k^2/T \to 0$ as k and $T \to \infty$, then*

$$d_2(\hat{F}_T, F) \to 0 \qquad \text{in probability.}$$

Recall that by Lemma 8.3 of Bickel and Freedman [2], convergence in the d_2 metric implies also convergence of second-order moments. Therefore, the above theorem implies $\hat{\boldsymbol{\Sigma}}_k \to \boldsymbol{\Sigma}$ which justifies asymptotically the use of the distribution $N(0, \hat{\boldsymbol{\Sigma}}_k)$ in generating the bootstrap errors $\boldsymbol{\varepsilon}_t^*$ in the case that a Gaussian distribution has been assumed for the true errors $\boldsymbol{\varepsilon}_t$.

The next result deals with the asymptotic properties of the bootstrap proposal in resampling the covariance structure of the process. Before formulating this result precisely, we fix some additional notation. Denote by $\mathbf{\Gamma}_k$ the $rk \times rk$ matrix defined by $\mathbf{\Gamma}_k = (\mathbf{\Gamma}(m-n))_{m,n=1,2,\ldots,k}$, where the (m, n)th block of this matrix is given by $\mathbf{\Gamma}(m-n) = E(\mathbf{y}_{t+n}\mathbf{y}_{t+m}^\top)$. Furthermore, let $\mathbf{Y}_{t,k}^*$ be the $kr \times 1$ vector defined by $\mathbf{Y}_{t,k}^* = (\mathbf{y}_t^{*\top}, \mathbf{y}_{t-1}^{*\top}, \ldots, \mathbf{y}_{t-k+1}^{*\top})^\top$ and denote by $\mathbf{\Gamma}_k^*$ the same matrix as $\mathbf{\Gamma}_k$ but obtained using the bootstrap process $\{\mathbf{y}_t^*, t \in \mathbf{Z}\}$, i.e., the (m, n)th block of this matrix is given by $\mathbf{\Gamma}^*(m-n) = E(\mathbf{y}_{t+n}^* \mathbf{y}_{t+m}^{*\top})$. Note that because $\mathbf{Y}_{t,k} = \sum_{j=0}^{\infty} (\mathbf{I}_k \otimes \mathbf{B}_j)\, \mathbf{U}_{t-j,k}$ and $\mathbf{Y}_{t,k}^* = \sum_{j=0}^{t+s-1} (\mathbf{I}_k \otimes \hat{\mathbf{B}}_{j,k})\, \mathbf{U}_{t-j,k}^*$, where $\mathbf{U}_{t,k} = (\boldsymbol{\varepsilon}_t^\top, \boldsymbol{\varepsilon}_{t-1}^\top, \ldots, \boldsymbol{\varepsilon}_{t-k+1}^\top)^\top$ and $\mathbf{U}_{t,k}^* = (\boldsymbol{\varepsilon}_t^{*\top}, \boldsymbol{\varepsilon}_{t-1}^{*\top}, \ldots, \boldsymbol{\varepsilon}_{t-k+1}^{*\top})^\top$, we have $\mathbf{\Gamma}_k = \sum_{i=0}^{\infty} \sum_{m=-k+1}^{k-1} (\mathbf{G}_m \otimes \mathbf{B}_i \mathbf{\Sigma} \mathbf{B}_{i-m}^\top)$ and $\mathbf{\Gamma}_k^* = \sum_{i=0}^{t+s-1} \sum_{m=-k+1}^{k-1} (\mathbf{G}_m \otimes \hat{\mathbf{B}}_{i,k} \mathbf{\Sigma}^* \hat{\mathbf{B}}_{i-m,k}^\top)$. In these expressions $\otimes$ is the Kronecker product, $\mathbf{\Sigma}^* = E(\boldsymbol{\varepsilon}_t^* \boldsymbol{\varepsilon}_t^{*\top})$ and $\mathbf{G}_m$ is the $k \times k$ matrix the (i, j)th element of which equals one if $j - i = m$ and zero otherwise.

THEOREM 2.5. *Let $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ satisfy* (1) *and assume that $k^3/T^{1/2} \to 0$ as $k, T \to \infty$. Then conditional on $\{\mathbf{y}_t, t = 1, 2, \ldots, T\}$*

(a) $\|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1 = o_P(1)$ *and*

(b) $\|\mathbf{\Gamma}_k^{*-1} - \mathbf{\Gamma}_k^{-1}\|_1 = o_P(1)$.

Note that from the assumptions imposed on the parameter matrices of the process $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ we have $\|\mathbf{\Gamma}(h)\| \to 0$ at a geometric rate as the lag $|h|$ increases. Losely speaking this means that the covariance structure of this process is well captured by the covariance matrices of the first few lags only. Thus by the above theorem and as T increases, the covariance structure of the bootstrap process becomes "similar" to the covariance structure of the true process. We expect, therefore, that the sampling behavior of the parameter estimates obtained from the observed part of the process can be well approximated by the sampling behavior of the parameter estimates obtained using the bootstrap observations.

3. BOOTSTRAPPING THE PARAMETER ESTIMATES

3.1. *Autoregressive Coefficients*

We first investigate the asymptotic properties of the bootstrap procedure applied in order to estimate the distribution of the statistic $(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}(k) - \mathbf{a}(k))$. Instead of $(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k))$, however, we consider in the following the statistic s_T^* which is defined by

$$s_T^* = (T-k)^{1/2}\, \mathbf{l}(k)^\top \operatorname{vec}\left[\left\{(T-k)^{-1} \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_{t+1}^* \mathbf{Y}_{t,k}^{*\top}\right\} \mathbf{\Gamma}_k^{*-1}\right]. \qquad (6)$$

To see the relation between these two statistics recall that $\hat{\mathbf{\Gamma}}^*_k = (T-k)^{-1} \sum_{t=k}^{T-1} \mathbf{Y}^*_{t,k} \mathbf{Y}^{*\top}_{t,k}$ and, therefore,

$$\begin{aligned}\hat{\mathbf{A}}^*(k) - \hat{\mathbf{A}}(k) &= (\hat{\mathbf{\Gamma}}^{*\top}_{1,k} - \hat{\mathbf{A}}(k)\, \hat{\mathbf{\Gamma}}^*_k)\, \hat{\mathbf{\Gamma}}^{*-1}_k \\ &= (T-k)^{-1} \left\{ \sum_{t=k}^{T-1} (\mathbf{y}^*_{t+1} - \hat{\mathbf{A}}(k)\, \mathbf{Y}^*_{t,k})\, \mathbf{Y}^{*\top}_{t,k} \right\} \hat{\mathbf{\Gamma}}^{*-1}_k \\ &= (T-k)^{-1} \left\{ \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}^*_{t+1} \mathbf{Y}^{*\top}_{t,k} \right\} \hat{\mathbf{\Gamma}}^{*-1}_k. \end{aligned} \tag{7}$$

Thus $(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k))$ and s^*_T differ by the fact that in s^*_T the theoretical matrix $\mathbf{\Gamma}^{*-1}_k$, instead of its estimator $\hat{\mathbf{\Gamma}}^{*-1}_k$, is used. The following theorem shows that these statistics have asymptotically the same behavior.

THEOREM 3.1. *Let $\{\mathbf{y}_t,\, t \in \mathbf{Z}\}$ satisfy* (1) *and assume that k is chosen as a function of T so that $k^{3.5}/T^{0.5} \to 0$ as k and $T \to \infty$. Then conditional on $\mathbf{y}_t$, $t = 1, 2, ..., T$,*

$$(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k)) = s^*_T + o_P(1) \qquad \textit{in probability.}$$

Now, let

$$s_T = (T-k)^{1/2}\, \mathbf{l}(k)^\top \operatorname{vec}\left[\left\{ (T-k)^{-1} \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_{t+1} \mathbf{Y}^\top_{t,k} \right\} \mathbf{\Gamma}^{-1}_k \right]. \tag{8}$$

Under the conditions mentioned in the previous section, Lewis and Reinsel [8] proved that $(T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}(k) - \mathbf{a}(k)) = s_T + o_P(1)$. Furthermore, they established $s_T \Rightarrow N(0, \mathbf{l}(k)^\top (\mathbf{\Gamma}^{-1}_k \otimes \mathbf{\Sigma})\, \mathbf{l}(k))$, where "$\Rightarrow$" stands for weak convergence. The next theorem shows that Mallow's distance between the conditional distribution of the bootstrap statistic s^*_T on one side and the statistic s_T on the other, goes to zero asymptotically.

THEOREM 3.2. *Let $\{\mathbf{y}_t,\, t \in \mathbf{Z}\}$ satisfy* (1). *Under the same conditions as in Theorem* 3.1 *we have*

$$d_2\{\mathscr{L}(s_T),\, \mathscr{L}(s^*_T \mid \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)\} \to 0 \qquad \textit{in probability.}$$

Now, by the above theorem and the fact that convergence in the d_2 metric implies also weak convergence of the corresponding random variables it follows, conditional on the observed part of the processes, that $s^*_T \Rightarrow N(0, \mathbf{l}(k)^\top (\mathbf{\Gamma}^{-1}_k \otimes \mathbf{\Sigma})\, \mathbf{l}(k))$ in probability. From this and from Theorem 3.1 above the following result is then obtained. It establishes the so-called asymptotic validity of the bootstrap proposal.

THEOREM 3.3. *Let* $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ *satisfy* (1). *Under the same conditions as in Theorem* 3.1

$$\mathscr{L}((T-k)^{1/2}\, \mathbf{l}(k)^\top (\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k)) \mid \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$$
$$\Rightarrow N(0, \mathbf{l}(k)^\top (\boldsymbol{\Gamma}_k^{-1} \otimes \boldsymbol{\Sigma})\, \mathbf{l}(k)) \qquad \textit{in probability.}$$

It is worth noting here that, although we restrict our discussion to the Yule–Walker estimators only, Theorem 3.3 above is also valid for other estimators of $\mathbf{A}(k)$, for instance, least squares estimators for which Theorem 2.1 holds.

It is well known that for any fixed $m \in \mathbf{N}$, the asymptotic distribution of $(T-k)^{1/2}\, \mathrm{vec}[(\hat{\mathbf{A}}_{1,k}, \hat{\mathbf{A}}_{2,k}, ..., \hat{\mathbf{A}}_{m,k}) - (\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_m)]$ is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V} \otimes \boldsymbol{\Sigma}$, where $\mathbf{V} = (\boldsymbol{\Gamma}_\infty^{ls})_{l,s=1,...,m}$ is the upper left $mr^2 \times mr^2$ corner of $\boldsymbol{\Gamma}_\infty^{-1}$ and $\boldsymbol{\Gamma}_\infty = (\boldsymbol{\Gamma}(i-j))_{i,j=1,2,...}$. One implication of Theorem 3.3 is then that conditional on $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T$,

$$(T-k)^{1/2}\, \mathrm{vec}[(\hat{\mathbf{A}}_{1k}^*, \hat{\mathbf{A}}_{2k}^*, ..., \hat{\mathbf{A}}_{mk}^*) - (\mathbf{A}_{1,k}, \hat{\mathbf{A}}_{2,k}, ..., \hat{\mathbf{A}}_{m,k})]$$
$$\Rightarrow N(\mathbf{0}, \mathbf{V} \otimes \boldsymbol{\Sigma}) \qquad \text{in probability.}$$

3.2. *Moving Average Coefficients*

Consider now the asymptotic behavior of the bootstrap procedure applied in order to estimate the sampling distribution of the estimated coefficient matrices in the moving average representation of the process.

THEOREM 3.4. *Let* $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ *satisfy* (1) *and assume that* k *is chosen as a function of* T *such that* $k^4/T^{1/2} \to 0$ *as* k *and* $T \to \infty$. *Then*

$$\mathscr{L}(\sqrt{T}\, \mathbf{l}(k)^\top (\hat{\mathbf{b}}^*(k) - \hat{\mathbf{b}}(k)) \mid \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$$
$$\Rightarrow N(0, \mathbf{l}(k)^\top \boldsymbol{\Omega}(k)\, \mathbf{l}(k)) \qquad \textit{in probability.}$$

where $\boldsymbol{\Omega}(k)$ *is the* $kr^2 \times kr^2$ *matrix given by* $\boldsymbol{\Omega}(k) = (\boldsymbol{\Sigma}^{-1} \otimes \sum_{j=0}^{k-1} \mathbf{B}_j \boldsymbol{\Sigma} \mathbf{B}_{m-n+j}^\top)_{n,m=1,2,...,k}$.

At the same lines as in proving Theorem 3.4 the following result which deals with the asymptotic distribution of the statistic $\sqrt{T}\, \mathbf{l}(k)^\top (\hat{\mathbf{b}}(k) - \mathbf{b}(k))$ can be established.

THEOREM 3.5. *Let* $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ *satisfy* (1) *and assume that* $k^3/T^{1/2} \to 0$ *as* k *and* $T \to \infty$. *Then*

$$\mathscr{L}(\sqrt{T}\, \mathbf{l}(k)^\top (\hat{\mathbf{b}}(k) - \mathbf{b}(k))) \Rightarrow N(\mathbf{0}, \mathbf{l}(k)^\top \boldsymbol{\Omega}(k)\, \mathbf{l}(k)),$$

where $\boldsymbol{\Omega}(k)$ *is given in Theorem* 3.4.

Note that the condition $k^3/T^{1/2} \to 0$ given in the above theorem differs from the one needed in establishing the asymptotic normality of the autoregressive parameter estimators. This condition is also somewhat slower than the one needed to establish the asymptotic normality if one is interested in the distribution of a fixed set of moving average coefficients only, cf. Lütkepohl [10, Chap. 9]. For this particular case of a fixed set of moving average coefficients, the following result appears. Its proof follows the same lines as the proof of Theorem 3.4.

THEOREM 3.6. *Let* $\{\mathbf{y}_t, t \in \mathbf{Z}\}$ *satisfy* (1). *If* $k^{3.5}/T^{0.5} \to 0$ *as* k *and* $T \to \infty$ *then for* $1 \leqslant m \leqslant k$ *fixed,*

$$\mathscr{L}(\sqrt{T}\,\mathrm{vec}[(\hat{\mathbf{B}}^*_{1,k}, \hat{\mathbf{B}}^*_{2,k}, ..., \hat{\mathbf{B}}^*_{m,k}) - (\hat{\mathbf{B}}_{1,k}, \hat{\mathbf{B}}_{2,k}, ..., \hat{\mathbf{B}}_{m,k})] \mid \mathbf{y}_1, ..., \mathbf{y}_T)$$
$$\Rightarrow N(\mathbf{0}, \mathbf{\Omega}(m))$$

in probability.

4. PROOFS

Proof of Theorem 2.2. Suppose that $\hat{\mathbf{B}}_k(z) = (\mathbf{I}_r - \sum_{j=1}^k \hat{\mathbf{A}}_{k,k} z^j)^{-1}$ exists. Write $\mathbf{B}(z) = (b_{ls}(z))_{l,s=1,...,r}$ and $\hat{\mathbf{B}}_k(z) = (\hat{b}_{ls,k}(z))_{l,s=1,...,r}$, where in this notation $b_{ls}(z)$ and $b_{ls,k}(z)$ are the polynomials in the (l, s)th position of $\mathbf{B}(z)$ and $\hat{\mathbf{B}}_k(z)$, i.e., $b_{ls}(z) = \sum_{j=0}^{\infty} b_{ls,j} z^j$, $\hat{b}_{ls,k}(z) = \sum_{i=0}^{\infty} \hat{b}_{ls,k,j} z^j$ and $b_{ls,0} = \hat{b}_{ls,k,0} = 1$ if $l = s$ and zero otherwise. Now, note that $\mathbf{B}(z) = (\det \mathbf{A}(z))^{-1} (a^+_{ls}(z))_{l,s=1,...,r}$ and $\hat{\mathbf{B}}_k(z) = (\det \hat{\mathbf{A}}_k(z))^{-1} (\hat{a}^+_{ls,k}(z))_{l,s=1,...,r}$, where the polynomials $a^+_{ls}(z)$ and $\hat{a}^+_{ls,k}(z)$ are the elements in the (l, s)th position of the adjoints of the matrices $\mathbf{A}(z) = \mathbf{I}_r - \sum_{j=1}^{\infty} \mathbf{A}_j z^j$ and $\hat{\mathbf{A}}_k(z) = \mathbf{I}_r - \sum_{j=1}^k \hat{\mathbf{A}}_{j,k} z^j$, respectively. Using this notation, the method used to evaluate the difference $|\hat{b}_{ls,k,j} - b_{ls,j}|$ follows the lines of proof of Lemma 1.4 in Kreiss [5]. Briefly, by assumption $\varepsilon > 0$ exists such that $\det(\mathbf{A}(z)) \neq 0$ for all $|z| \leqslant 1 + \varepsilon$. Therefore, using Theorem 2.1 we have after some simple manipulations that for all $|z| \leqslant 1 + k^{-1}$, $\|\hat{\mathbf{A}}_k(z) - \mathbf{A}(z)\| \leqslant O_P(k/T^{1/2}) \to 0$. Thus for large T, $\det(\hat{\mathbf{A}}_k(z)) \neq 0$ for all $|z| \leqslant 1 + k^{-1}$. Furthermore, for $l, s = 1, 2, ..., r$, we have $|\hat{a}_{ls,k}(z) - a_{ls}(z)| = O_P(k/T^{1/2})$, where $\hat{a}_{ls,k}(z)$ and $a_{ls}(z)$ denote the polynomials in the (l, s)th position of the matrices $\hat{\mathbf{A}}_k(z)$ and $\mathbf{A}(z)$. Therefore, $|\det(\hat{\mathbf{A}}_k(z)) - \det(\mathbf{A}(z))|$ and $|\hat{a}^+_{ls,k}(z) - a^+_{ls}(z)|$ are both $O_P(k/T^{1/2})$. Now,

$$|\hat{b}_{ls,k,j} - b_{ls,j}| = (j!)^{-1} \left|(\hat{b}^{(j)}_{ls,k}(z) - b^{(j)}_{ls}(z))|_{z=0}\right|,$$

where $\hat{b}^{(j)}_{ls,k}(z)|_{z=z_0}$ and $b^{(j)}_{ls}(z)|_{z=z_0}$ denote the jth derivatives at $z = z_0$ of $\hat{b}_{ls,k}(z)$ and $b_{ls}(z)$, respectively. Using Cauchy's inequality for the derivatives of an analytic function (cf. Churchill and Brown [4]), we get then

$$
\begin{aligned}
|\hat{b}_{ls,k,j} - b_{ls,j}| &\leqslant (1+k^{-1})^{-j} \max_{z=1+k^{-1}} |\hat{b}_{ls,k}(z) - b_{ls}(z)| \\
&= (1+k^{-1})^{-j} \max_{z=1+k^{-1}} \\
&\quad \times |(\det \hat{\mathbf{A}}_k(z))^{-1} \hat{a}^{+}_{ls,k}(z) - (\det \mathbf{A}(z))^{-1} a^{+}_{ls}(z)| \\
&\leqslant \text{const}(1+k^{-1})^{-j} \frac{k}{T^{1/2}}.
\end{aligned}
$$

Proof of Theorem 2.3. We have

$$
\begin{aligned}
E\,\|\mathbf{w}_1 - \mathbf{w}_2\|^2 &= (T-k)^{-1} \sum_{t=k+1}^{T} \|(\tilde{\boldsymbol{\varepsilon}}_{k,t} - \bar{\boldsymbol{\varepsilon}}_k) - \boldsymbol{\varepsilon}_t\|^2 \\
&\leqslant 2(T-k)^{-1} \sum_{t=k+1}^{T} \|\boldsymbol{\varepsilon}_t - \tilde{\boldsymbol{\varepsilon}}_{k,t}\|^2 + 2\,\|\bar{\boldsymbol{\varepsilon}}_k\|^2 \\
&\leqslant 4(T-k)^{-1} \sum_{t=k+1}^{T} \\
&\quad \times \left\{ \|(\hat{\mathbf{A}}(k) - \mathbf{A}(k))\,\mathbf{Y}_{t,k}\|^2 + \left\| \sum_{j=k+1}^{\infty} \mathbf{A}_j \mathbf{y}_{t-j} \right\|^2 \right\} + 2\,\|\bar{\boldsymbol{\varepsilon}}_k\|^2 \\
&= O_P(k^2/T) + O_P(1) \sum_{j=k+1}^{\infty} \|\mathbf{A}_j\|^2 + O_P(1/T).
\end{aligned}
$$

The last equality has been obtained using $\|\mathbf{Y}_{t,k}\|^2 = O_P(k)$ and the assertion of Theorem 2.1.

Proof of Theorem 2.5. Recall first that $\|\mathbf{C}\|_1 \leqslant \|\mathbf{C}\|$. Now, for $h \in \{-k+1, -k+2, ..., k-2, k-1\}$ we have $\boldsymbol{\Gamma}(h) = \sum_{j=0}^{\infty} \mathbf{B}_j \boldsymbol{\Sigma} \mathbf{B}_{j+h}$ and $\hat{\boldsymbol{\Gamma}}(h) = \sum_{j=0}^{t+s-1} \hat{\mathbf{B}}_{j,k} \boldsymbol{\Sigma}^* \hat{\mathbf{B}}_{j+h,k}$. Furthermore, $\sum_{j=t+s}^{\infty} \|\mathbf{B}_j \boldsymbol{\Sigma} \mathbf{B}_{j+h}\| = o(1)$ for $t \to \infty$. Thus,

$$
\begin{aligned}
\|\boldsymbol{\Gamma}^*(h) - \boldsymbol{\Gamma}(h)\| &\leqslant \sum_{j=0}^{t+s-1} \|\hat{\mathbf{B}}_{j,k} \boldsymbol{\Sigma}^* \hat{\mathbf{B}}^{\top}_{j+h,k} - \mathbf{B}_j \boldsymbol{\Sigma} \mathbf{B}_{j+h}\| + o(1) \\
&\leqslant \sum_{j=0}^{t+s-1} \Big\{ \|\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j\|\ \|\boldsymbol{\Sigma}^*\|\ \|\hat{\mathbf{B}}_{j+h,k}\| \\
&\quad + \|\mathbf{B}_j\|\ \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*\|\ \|\hat{\mathbf{B}}_{j+h,k}\| \\
&\quad + \sum_{j=0}^{t+s-1} \|\mathbf{B}_j\|\ \|\boldsymbol{\Sigma}\|\ \|\hat{\mathbf{B}}_{j+h,k} - \mathbf{B}_{j+h}\| \Big\} + o(1).
\end{aligned}
\tag{9}
$$

Now, for $k^2/T^{1/2} \to 0$ all three terms on the right-hand side of the last inequality go to zero. To see this consider the first term. From Theorem 2.2 we have that $\rho \in (0, 1)$ exists such that

$$\|\hat{\mathbf{B}}_{j,k}\| \leqslant C[\rho^j + (1+k^{-1})^{-j} kT^{-1/2}]. \tag{10}$$

Furthermore, because of $\sum_{j=0}^{t+s-1} (k/(k+1))^j \leqslant \sum_{j=0}^{\infty} (k/(k+1))^j \leqslant k$ we get that the first term is $O_P(k^2/T^{1/2})$. Using a similar argument it can be seen that the third term in (9) is $O_P(k^2/T^{1/2})$ too. Finally, using the notation $\mathbf{\Sigma}_T = (T-k)^{-1} \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top$, we have for the second term

$$\begin{aligned}
\|\mathbf{\Sigma}^* - \mathbf{\Sigma}\| &= \left\|(T-k)^{-1} \sum_{t=k}^{T-1} \hat{\boldsymbol{\varepsilon}}_{k,t} \hat{\boldsymbol{\varepsilon}}_{k,t}^\top - E\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right\| \\
&\leqslant (T-k)^{-1} \sum_{t=k}^{T-1} \|\hat{\boldsymbol{\varepsilon}}_{k,t} \hat{\boldsymbol{\varepsilon}}_{k,t}^\top - \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\| + \|\mathbf{\Sigma}_T - \mathbf{\Sigma}\| \\
&\leqslant (T-k)^{-1} \sum_{t=k}^{T-1} \|\hat{\boldsymbol{\varepsilon}}_{k,t} - \boldsymbol{\varepsilon}_t\| \, \|\hat{\boldsymbol{\varepsilon}}_{k,t}\| \\
&\quad + (T-k)^{-1} \sum_{t=k}^{T-1} \|\boldsymbol{\varepsilon}_t\| \, \|\hat{\boldsymbol{\varepsilon}}_{k,t} - \boldsymbol{\varepsilon}_t\| + o(1).
\end{aligned}$$

Note that the last inequality has been obtained using the property $\|\mathbf{u}\mathbf{v}^\top\| = \|\mathbf{u}\| \, \|\mathbf{v}\|$ for two vectors $\mathbf{u}$ and $\mathbf{v}$, and $\|\mathbf{\Sigma}_T - \mathbf{\Sigma}\| \to 0$. Now, substituting for $\hat{\boldsymbol{\varepsilon}}_{k,t}$ and $\boldsymbol{\varepsilon}_t$ and at the same lines as in the proof of Theorem 2.3, we get for the first two terms on the right-hand side of the last inequality above that they are $O_P(k/T^{1/2})$. From this and by an argument similar to those used for the first term in (9) we get that $\sum_{j=0}^{t+s-1} \|\mathbf{B}_j\| \, \|\mathbf{\Sigma} - \mathbf{\Sigma}^*\| \, \|\hat{\mathbf{B}}_{j+h,k}\| = O_P(k/T^{1/2})$. This proves that $\|\mathbf{\Gamma}^*(h) - \mathbf{\Gamma}(h)\| = O_P(k^2/T^{1/2})$ and because $\|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1^2 \leqslant \|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|^2$ we conclude that $\|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1^2 = O_P(k^6/T)$.

Part (b) of the theorem is proved using part (a) and an argument similar to that applied by Berk [1, p. 493]. In particular, recall that $\|\mathbf{\Gamma}_k^{-1}\|_1$ is uniformly bounded above by a positive constant $\mathscr{F}$ for all k. We have then

$$\begin{aligned}
\|\mathbf{\Gamma}_k^{*\,-1} - \mathbf{\Gamma}_k^{-1}\|_1 &= \|\mathbf{\Gamma}_k^{*\,-1}(\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k)\, \mathbf{\Gamma}_k^{-1}\|_1 \\
&\leqslant \|\mathbf{\Gamma}_k^{*\,-1}\|_1 \, \|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1 \, \|\mathbf{\Gamma}_k^{-1}\|_1 \\
&\leqslant (\mathscr{F} + \|\mathbf{\Gamma}_k^{*\,-1} - \mathbf{\Gamma}_k^{-1}\|_1)\, \|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1 \, \mathscr{F}.
\end{aligned}$$

Therefore, and by the assumption $k^3/T^{0,5} \to 0$, we get

$$0 \leqslant \frac{\|\mathbf{\Gamma}_k^{*\,-1} - \mathbf{\Gamma}_k^{-1}\|}{(\mathscr{F} + \|\mathbf{\Gamma}_k^{*\,-1} - \mathbf{\Gamma}_k^{-1}\|_1)\, \mathscr{F}} \leqslant \|\mathbf{\Gamma}_k^* - \mathbf{\Gamma}_k\|_1 \to 0.$$

Proof of Theorem 3.1. Let $\hat{s}_T^* = (T-k)^{1/2}\,\mathbf{l}(k)^\top \operatorname{vec}[\{(T-k)^{-1}\ \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_{t+1}^*\, \mathbf{Y}_{t,k}^{*\top}\}\ \hat{\boldsymbol{\Gamma}}_k^{*\,-1}]$ and note that $\hat{s}_T^* = (T-k)^{1/2}\,\mathbf{l}(k)^\top\,(\hat{\mathbf{a}}(k)-\mathbf{a}(k))$. We then have

$$
\begin{aligned}
|\hat{s}_T^* - s_T^*| &= \left|(T-k)^{1/2}\,\mathbf{l}(k)^\top \operatorname{vec}\left[\left\{(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\}(\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1})\right]\right|\\
&\leqslant (T-k)^{1/2}\,\|\mathbf{l}(k)\|\\
&\quad\times\left\|\operatorname{vec}\left[\left\{(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\}(\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1})\right]\right\|\\
&\leqslant M_2^{1/2}(T-k)^{1/2}\left\|(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\|\,\|\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1}\|_1.
\end{aligned}
$$

In obtaining the last inequality we made use of $\|\operatorname{vec}(\mathbf{AB})\| = \|\mathbf{AB}\|$ and of the inequality $\|\mathbf{AB}\| \leqslant \|\mathbf{A}\|\,\|\mathbf{B}\|_1$, relating the matrix norms $\|\cdot\|$ and $\|\cdot\|_1$. From the independence between $\boldsymbol{\varepsilon}_{t+1}^*$ and $\mathbf{Y}_{t,k}^*$ we get

$$
\begin{aligned}
E\left\|(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\|^2 &= (T-k)^{-2}\sum_{t=k}^{T-1}E(\boldsymbol{\varepsilon}_{t+1}^{*\top}\boldsymbol{\varepsilon}_{t+1}^*)\,E(\mathbf{Y}_{t,k}^{*\top}\mathbf{Y}_{t,k}^*)\\
&= k(T-k)^{-1}\operatorname{tr}(\boldsymbol{\Sigma}^*)\operatorname{tr}(\boldsymbol{\Gamma}^*(0)).
\end{aligned}
$$

Therefore, $|\hat{s}_T^* - s_T^*| = O_P(1)\,k^{1/2}\,\|\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1}\|_1$ and it remains to show that $k^{1/2}\,\|\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1}\|_1 \to 0$. Now,

$$
k^{1/2}\,\|\hat{\boldsymbol{\Gamma}}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{*\,-1}\|_1 \leqslant k^{1/2}\,\|\boldsymbol{\Gamma}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{-1}\|_1 + k^{1/2}\,\|\boldsymbol{\Gamma}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{-1}\|_1 \tag{11}
$$

and by Theorem 2.5 we need only show that the first term on the right-hand side of (11) goes to zero. However, by the same arguments as in the proof of Theorem 2.5, part (b), it suffices to show that $k^{1/2}\,\|\hat{\boldsymbol{\Gamma}}_k^* - \boldsymbol{\Gamma}_k\| \to 0$. To prove the last statement we use for $h \in \{-k+1, -k+2, ..., k-1\}$ the expression

$$
\begin{aligned}
\hat{\boldsymbol{\Gamma}}^*(h) - \boldsymbol{\Gamma}(h) &= \sum_{j=0}^{t+s-1}\left(\hat{\mathbf{B}}_{j,k}\left((T-k)^{-1}\sum_{t=1}^{T-1}\boldsymbol{\varepsilon}_{t-j}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}\right)\hat{\mathbf{B}}_{j+h,k}^\top - \mathbf{B}_j\boldsymbol{\Sigma}\mathbf{B}_{j+h}^\top\right)\\
&\quad+\sum_{i=0}^{t+s-1}\sum_{\substack{j=0\\ j\neq i}}^{t+s-1}\hat{\mathbf{B}}_{i,k}\left((T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t-i}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}\right)\hat{\mathbf{B}}_{j+h,k}^\top + o(1)\\
&= T_1 + T_2 + o(1),
\end{aligned}
$$

where the $o(1)$ term is due to the asymptotically negligible term $\sum_{j=t+s}^{\infty}\mathbf{B}_j\boldsymbol{\Sigma}\mathbf{B}_{j+h}$. Now, $\{\boldsymbol{\varepsilon}_t^*, t = -s+1, -s+2, ..., T\}$ is a seqence of independent random vectors distributed according to $\hat{F}_T$ with $E\boldsymbol{\varepsilon}_t^* = 0$

and $E\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*\top}=\boldsymbol{\Sigma}^*\to\boldsymbol{\Sigma}$ for $k^2/T\to 0$ as $k, T\to\infty$. Therefore for $i\neq j$, $(T-k)^{-1/2}\sum_{t=k}^{T-k}\boldsymbol{\varepsilon}_{t-i}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}=O_P(1)$ and using (10) it is easily seen that

$$T_2=\sum_{i=0}^{t+s-1}\sum_{\substack{j=0\\ j\neq i}}^{t+s-1}\hat{\mathbf{B}}_{i,k}\left((T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t-i}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}\right)\hat{\mathbf{B}}_{j+h,k}^{\top}=O_P((T-k)^{-1/2}).$$

For the term T_1 we have

$$\begin{aligned}
\|T_1\| &\leqslant \sum_{j=0}^{t+s-1}\left\|\hat{\mathbf{B}}_{j,k}(T-k)^{-1}\sum_{t=1}^{T-1}\boldsymbol{\varepsilon}_{t-j}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}\hat{\mathbf{B}}_{j+h,k}^{\top}-\mathbf{B}_j\boldsymbol{\Sigma}\mathbf{B}_{j+h}^{\top}\right\|\\
&\leqslant \sum_{j=0}^{t+s-1}\|\hat{\mathbf{B}}_{j,k}-\mathbf{B}_j\|\,\|\boldsymbol{\Sigma}\|\,\|\hat{\mathbf{B}}_{j+h,k}\|\\
&\quad+\sum_{j=0}^{t+s-1}\|\hat{\mathbf{B}}_{j,k}\|\left\|(T-k)^{-1}\sum_{t=1}^{T-1}\boldsymbol{\varepsilon}_{t-j}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}-\boldsymbol{\Sigma}\right\|\,\|\hat{\mathbf{B}}_{j+h,k}\|\\
&\quad+\sum_{j=0}^{t+s-1}\|\mathbf{B}_j\|\,\|\boldsymbol{\Sigma}\|\,\|\hat{\mathbf{B}}_{j+h,\,k}-\mathbf{B}_{j+h}\|\\
&=S_1+S_2+S_3.
\end{aligned}$$

At the same lines as in the proof of Theorem 2.5, part (a), we get that S_1 and S_3 are $O_P(k^2/T^{1/2})$. Furthermore, because

$$\begin{aligned}
\left\|(T-k)^{-1}\sum_{t=1}^{T-1}\boldsymbol{\varepsilon}_{t-j}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}-\boldsymbol{\Sigma}\right\| &\leqslant \left\|(T-k)^{-1}\sum_{t=1}^{T-1}\boldsymbol{\varepsilon}_{t-j}^*\boldsymbol{\varepsilon}_{t-j}^{*\top}-\boldsymbol{\Sigma}^*\right\|+\|\boldsymbol{\Sigma}^*-\boldsymbol{\Sigma}\|\\
&=O_P((T-k)^{-1/2})+O_P(k/T^{1/2})
\end{aligned}$$

we conclude using (10) that $S_2=O_P(k/T^{1/2})$. Thus $\|\hat{\boldsymbol{\Gamma}}_k^*-\boldsymbol{\Gamma}_k\|^2=O_P(k^6/T)$ and $k^{1/2}\,\|\hat{\boldsymbol{\Gamma}}_k^*-\boldsymbol{\Gamma}_k\|\to 0$ by the assumption $k^{3.5}/T^{0.5}\to 0$.

Proof of Theorem 3.2. Substituting for s_T and s_T^* and using the definition of the d_2 metric it suffices to show that

$$\begin{aligned}
E\Bigg|(T-k)^{1/2}\,\mathbf{l}(\mathrm{k})^{\top}\operatorname{vec}\Bigg[&\left\{(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^{\top}\right\}\boldsymbol{\Gamma}_k^{-1}\\
&-\left\{(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t,k}^*\mathbf{Y}_{t,u}^{*\top}\right\}\boldsymbol{\Gamma}_k^{*\,-1}\Bigg]\Bigg|^2\\
\leqslant 2S_1+2S_2\to 0,&
\end{aligned}$$

where

$$S_1=E\left|(T-k)^{1/2}\,\mathbf{l}(k)^{\top}\operatorname{vec}\left[\left\{(T-k)^{-1}\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\}(\boldsymbol{\Gamma}_k^{*\,-1}-\boldsymbol{\Gamma}_k^{-1})\right]\right|^2 \tag{12}$$

and

$$S_2 = E\left|(T-k)^{1/2}\,\mathbf{l}(k)^\top \operatorname{vec}\left[\left\{(T-k)^{-1}\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^\top - \boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top})\right\}\boldsymbol{\Gamma}_k^{-1}\right]\right|^2. \tag{13}$$

Consider first the term S_1. Because of the independence between $\boldsymbol{\varepsilon}_{t+1}^*$ and $\mathbf{Y}_{t,k}^*$, we have

$$\begin{aligned}S_1 &= (T-k)^{-1}\,\|\mathbf{l}(k)\|^2\,E\left\|\operatorname{vec}\left(\left\{\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\}(\boldsymbol{\Gamma}_k^{*-1}-\boldsymbol{\Gamma}_k^{-1})\right)\right\|^2\\
&\leqslant M_2(T-k)^{-1}\,E\left\|\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top}\right\|^2\,\|\boldsymbol{\Gamma}_k^{*-1}-\boldsymbol{\Gamma}_k^{-1}\|_1^2\\
&= M_2\operatorname{tr}(\boldsymbol{\Sigma}^*)\operatorname{tr}(\boldsymbol{\Gamma}^*(0))\,k\,\|\boldsymbol{\Gamma}_k^{*-1}-\boldsymbol{\Gamma}_k^{-1}\|_1^2 \to 0\end{aligned}$$

by part (b) of Theorem 2.5 and the assumption $k^{3.5}/T^{0.5}\to 0$.

Consider next the term S_2. For this term we have

$$\begin{aligned}&E\left|(T-k)^{1/2}\,\mathbf{l}(k)^\top \operatorname{vec}\left[\left\{(T-k)^{-1}\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^\top - \boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top})\right\}\boldsymbol{\Gamma}_k^{-1}\right]\right|^2\\
&\quad\leqslant (T-k)^{-1}\,\|\mathbf{l}(k)\|^2\,E\left\|\operatorname{vec}\left\{\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^\top - \boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top})\right\}\boldsymbol{\Gamma}_k^{-1}\right\|^2\\
&\quad\leqslant M_2\,\|\boldsymbol{\Gamma}_k^{-1}\|_1^2(T-k)^{-1}\,E\left\|\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^\top - \boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top})\right\|^2.\end{aligned}$$

Now, rewrite the term following $\|\boldsymbol{\Gamma}_k^{-1}\|_1^2$ in the above inequality as

$$\begin{aligned}&(T-k)^{-1}E\left\|\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}\mathbf{Y}_{t,k}^\top - \boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top})\right\|^2\\
&\quad\leqslant 2(T-k)^{-1}\,E\left\|\sum_{t=k}^{T-1}\boldsymbol{\varepsilon}_{t+1}(\tilde{\mathbf{Y}}_{t,k}-\mathbf{Y}_{t,k})^\top\right\|^2\\
&\qquad + 2(T-k)^{-1}\,E\left\|\sum_{t=k}^{T-1}(\boldsymbol{\varepsilon}_{t+1}^*\mathbf{Y}_{t,k}^{*\top} - \boldsymbol{\varepsilon}_{t+1}\tilde{\mathbf{Y}}_{t,k}^\top)\right\|^2,\end{aligned} \tag{14}$$

where $\tilde{\mathbf{Y}}_{t,k} = (\tilde{\mathbf{y}}_t^\top, \tilde{\mathbf{y}}_{t-1}^\top, \ldots, \tilde{\mathbf{y}}_{t-k+1}^\top)^\top$ and $\tilde{\mathbf{y}}_t$ is the process defined by $\tilde{\mathbf{y}}_t = \sum_{j=0}^{t+s-1}\hat{\mathbf{B}}_{j,k}\boldsymbol{\varepsilon}_{t-j}$. Confirm now that for any positive definite matrix $\mathbf{S}$ we have $\operatorname{tr}(\mathbf{A}^\top\mathbf{A}\mathbf{S}) \leqslant \|\mathbf{A}\|^2\operatorname{tr}(\mathbf{S})$, where $\mathbf{A}$ is an appropriate matrix. Using this property and

$$\tilde{\mathbf{Y}}_{t,k} - \mathbf{Y}_{t,k} = \sum_{j=0}^{t+s-1}\{\mathbf{I}_k\otimes(\hat{\mathbf{B}}_{j,k}-\mathbf{B}_j)\}\,\mathbf{U}_{t-j,k} - \sum_{j=t+s}^{\infty}(\mathbf{I}_k\otimes\mathbf{B}_j)\,\mathbf{U}_{t-j,k}$$

we get for the first term on the right-hand side of (14) conditional on $\mathbf{y}_1, ..., \mathbf{y}_T$

$$(T-k)^{-1} E \left\| \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_{t+1} (\tilde{\mathbf{Y}}_{t,k} - \mathbf{Y}_{t,k})^\top \right\|^2$$

$$= (T-k)^{-1} \sum_{t=k}^{T-1} E \, \|\boldsymbol{\varepsilon}_{t+1}\|^2 \, E \, \|\tilde{\mathbf{Y}}_{t,k} - \mathbf{Y}_{t,k}\|^2$$

$$\leqslant 2(T-k)^{-1} \operatorname{tr}(\boldsymbol{\Sigma}) \sum_{t=k}^{T-1}$$

$$\times E \left\{ \left\| \sum_{j=0}^{t+s-1} \{\mathbf{I}_k \otimes (\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j)\} \, \mathbf{U}_{t-j,k} \right\|^2 + \left\| \sum_{j=t+s}^{\infty} (\mathbf{I}_k \otimes \mathbf{B}_j) \, \mathbf{U}_{t-j,k} \right\|^2 \right\}$$

$$= 2k(T-k)^{-1} \operatorname{tr}(\boldsymbol{\Sigma})$$

$$\times \sum_{t=k}^{T-1} \left\{ \sum_{j=0}^{t+s-1} \operatorname{tr}\{(\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j)^\top (\hat{\mathbf{B}}_{j,k} - \mathbf{B}_j) \, \boldsymbol{\Sigma}\} + \sum_{j=t+s}^{\infty} \operatorname{tr}\{\mathbf{B}_j^\top \mathbf{B}_j \boldsymbol{\Sigma})\right\}$$

$$= O_P(k^4/T).$$

For the second term on the right-hand side of (14) we have

$$(T-k)^{-1} E \left\| \sum_{t=k}^{T-1} (\boldsymbol{\varepsilon}^*_{t+1} \mathbf{Y}^{*\top}_{t,k} - \boldsymbol{\varepsilon}_{t+1} \tilde{\mathbf{Y}}^\top_{t,k}) \right\|^2$$

$$\leqslant 2(T-k)^{-1} \sum_{t=k}^{T-1} \{E \, \|\boldsymbol{\varepsilon}^*_{t+1}\|^2 \, E \, \|\mathbf{Y}^*_{t,k} - \tilde{\mathbf{Y}}_{t,k}\|^2 + E \, \|\boldsymbol{\varepsilon}^*_{t+1} - \boldsymbol{\varepsilon}_{t+1}\|^2 \, E \, \|\tilde{\mathbf{Y}}_{t,k}\|^2\}.$$

For the last term on the right-hand side of the last inequality, we have at the same lines as in the proof of Theorem 2.3 and using $E \, \|\tilde{\mathbf{Y}}_{t,k}\|^2 = k \operatorname{tr}(\tilde{\boldsymbol{\Gamma}}(0))$, where $\tilde{\boldsymbol{\Gamma}}(0) = E(\tilde{\mathbf{y}}_t \mathbf{y}_t^\top)$, that

$$(T-k)^{-1} \sum_{t=k}^{T-1} E \, \|\boldsymbol{\varepsilon}^*_{t+1} - \boldsymbol{\varepsilon}_{t+1}\|^2 \, E \, \|\tilde{\mathbf{Y}}_{t,k}\|^2 = O_P(k^3/T).$$

Finally, after some straightforward manipulations it can be seen that $(T-k)^{-1} \sum_{t=k}^{T-1} E \, \|\boldsymbol{\varepsilon}^*_{t+1}\|^2 \, E \, \|\mathbf{Y}^*_{t,k} - \tilde{\mathbf{Y}}_{t,k}\|^2 = O_P(k^3/T)$.

In proving Theorem 3.4 we make use of the following intermediate result.

THEOREM 4.1. *Let $k^3/T^{1/2} \to 0$ and suppose that the infinite series expansions $(\mathbf{I}_r - \sum_{j=1}^{k} \hat{\mathbf{A}}_{j,k} z^j)^{-1}$ and $(\mathbf{I}_r - \sum_{j=1}^{k} \hat{\mathbf{A}}^*_{j,k} z^j)^{-1}$ exists. Furthermore, denote by $\{\hat{\mathbf{B}}^*_{j,k}, j \in \mathbf{N}\}_{k \in \mathbf{N}}$ the sequence of the coefficient matrices appearing* *in the last expansion. Then a constant $C \in (0, \infty)$ exists such that conditional on $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T$ and for all $j \in \mathbf{N}$,*

$$\|\widehat{\mathbf{B}}^*_{j,k} - \widehat{\mathbf{B}}_{j,k}\| \leqslant C(1+k^{-1})^{-j} \frac{k^2}{T^{1/2}}.$$

Proof. Consider first $\|\widehat{\mathbf{A}}^*(k) - \widehat{\mathbf{A}}(k)\|$. From (7) this term is bounded by $\|\widehat{\mathbf{A}}^*(k) - \widehat{\mathbf{A}}(k)\| \leqslant \{\|U_1\| + \|U_2\|\}\ \|\widehat{\mathbf{\Gamma}}^{*\,-1}_k\|_1$, where

$$U_1 = (T-k)^{-1} \sum_{t=k}^{T-1} (\boldsymbol{\varepsilon}^*_{t+1} - \boldsymbol{\varepsilon}_{t+1})\, \mathbf{Y}^{*\top}_{t,k}, \qquad U_2 = (T-k)^{-1} \sum_{t=k}^{T-1} \boldsymbol{\varepsilon}_{t+1} \mathbf{Y}^{*\top}_{t,k}.$$

Using similar arguments to those applied in the proof of Theorems 3.2 and 3.1 it can be seen that $E\|U_1\|^2 = O_P(k^3/T)$ and $E\|U_2\|^2 = O_P(k/T-k))$. Furthermore, because $\|\widehat{\mathbf{\Gamma}}^{*\,-1}_k\|_1 \leqslant \|\mathbf{\Gamma}^{-1}_k\|_1 + \|\widehat{\mathbf{\Gamma}}^{*\,-1}_k - \mathbf{\Gamma}^{-1}_k\|_1$ we have that $\|\widehat{\mathbf{\Gamma}}^{*\,-1}_k\|_1 = O_P(1)$ for $k^3/T^{0.5} \to 0$. Therefore, $\|\widehat{\mathbf{A}}^*(k) - \widehat{\mathbf{A}}(k)\| = O_P(k^{3/2}/T^{1/2})$. The assertion of the theorem follows then by applying the same arguments as in the proof of Theorem 2.2.

Proof of Theorem 3.4. Using $\widehat{\mathbf{B}}^*_{j,k} = \mathbf{J}_k \widehat{\mathbf{\Pi}}^{*\,j}_k \mathbf{J}^\top_k$ and $\widehat{\mathbf{B}}_{j,k} - \mathbf{J}_k \widehat{\mathbf{\Pi}}^j_k \mathbf{J}^\top_k$ we have

$$\begin{aligned}
\operatorname{vec}(\widehat{\mathbf{B}}^*_{j,k} - \widehat{\mathbf{B}}_{j,k}) &= \operatorname{vec}\{\mathbf{J}_k(\widehat{\mathbf{\Pi}}^{*\,j}_k - \widehat{\mathbf{\Pi}}^j_k)\, \mathbf{J}^\top_k\} \\
&= \operatorname{vec}\left\{\sum_{i=0}^{j-1} \mathbf{J}_k \widehat{\mathbf{\Pi}}^{*\,i}_k (\widehat{\mathbf{\Pi}}^*_k - \widehat{\mathbf{\Pi}}_k)\, \widehat{\mathbf{\Pi}}^{j-i-1}_k \mathbf{J}^\top_k\right\} \\
&= \sum_{i=0}^{j-1} (\mathbf{J}_k \widehat{\mathbf{\Pi}}^{j-1-i^\top}_k \otimes \widehat{\mathbf{B}}^*_{i,k})(\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k));
\end{aligned}$$

cf. Schmidt [13]. Thus,

$$\sqrt{T}\, \mathbf{l}(k)^\top (\hat{\mathbf{b}}^*(k) - \hat{\mathbf{b}}(k)) = \sqrt{T}\, \mathbf{l}(k)^\top (\widehat{\mathbf{\Psi}}_1(k) + \widehat{\mathbf{\Psi}}^*_2(k))(\hat{\mathbf{a}}^*(k) - \hat{\mathbf{a}}(k)), \qquad (15)$$

where the $kr^2 \times kr^2$ matrices $\widehat{\mathbf{\Psi}}_1(k)$ and $\widehat{\mathbf{\Psi}}^*_2(k)$ are given by

$$\widehat{\mathbf{\Psi}}_1(k) = \begin{pmatrix} \mathbf{J}_k \otimes \mathbf{I}_r \\ \sum_{j=0}^{1} \mathbf{J}_k \mathbf{\Pi}^{\top 1-j}_k \otimes \widehat{\mathbf{B}}_{j,k} \\ \vdots \\ \sum_{j=0}^{k-1} \mathbf{J}_k \widehat{\mathbf{\Pi}}^{\top k-j-1}_k \otimes \widehat{\mathbf{B}}_{j,k} \end{pmatrix},$$

$$\mathbf{\Psi}^*_2(k) = \begin{pmatrix} \mathbf{J}_k \otimes (\widehat{\mathbf{B}}^*_{0,k} - \widehat{\mathbf{B}}_{0,k}) \\ \sum_{j=0}^{1} \mathbf{J}_k \widehat{\mathbf{\Pi}}^{\top 1-j}_k \otimes (\widehat{\mathbf{B}}^*_{j,k} - \widehat{\mathbf{B}}_{j,k}) \\ \vdots \\ \sum_{j=0}^{k-1} \mathbf{J}_k \mathbf{\Pi}^{\top k-j-1}_k \otimes (\widehat{\mathbf{B}}^*_{j,k} - \widehat{\mathbf{B}}_{j,k}) \end{pmatrix}.$$

We first show that, under the assumptions of the theorem,

$$\sqrt{T}\,\mathbf{l}(k)^\top\,\hat{\mathbf{\Psi}}_2^*(k)\,(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k)) \Rightarrow 0 \qquad \text{in probability.} \tag{16}$$

This can be proved by partioning the vector $\mathbf{l}(k)$ as follows: $\mathbf{l}(k)=(\mathbf{l}_1^\top, \mathbf{l}_2^\top, ..., \mathbf{l}_k^\top)^\top$, where $\mathbf{l}_i=(\mathbf{l}_{1,i}^\top, \mathbf{l}_{2,i}^\top, ..., \mathbf{l}_{r,i}^\top)^\top$ and each of the $\mathbf{l}_{s,i}$, $s=1, 2, ..., r$ is a $r\times 1$ vector. We have then

$$\begin{aligned}
&|\sqrt{T}\,\mathbf{l}(k)^\top\,\hat{\mathbf{\Psi}}_2^*(k)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))|\\
&\quad=\Bigg|\sum_{h=0}^{k-1}\sum_{j=0}^{h}\operatorname{vec}(\hat{\mathbf{B}}_{j,k}^*-\hat{\mathbf{B}}_{j,k})^\top\sqrt{T}\,(\mathbf{I}_r\otimes\mathbf{l}_{1,h+1}, \mathbf{I}_r\otimes\mathbf{l}_{2,h+1}, ..., \mathbf{I}_r\otimes\mathbf{l}_{r,h+1})\\
&\qquad\times(\mathbf{J}_k\hat{\mathbf{\Pi}}_k^{\top h-j}\otimes\mathbf{I}_r)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))\Bigg|.
\end{aligned}\tag{17}$$

Now, each component of the $r^2\times 1$ vector

$$\begin{aligned}
\hat{\mathbf{c}}_{h-j}(k)&=\sqrt{T}\,(\mathbf{I}_r\otimes\mathbf{l}_{1,h+1}, \mathbf{I}_r\otimes\mathbf{l}_{2,h+1}, ..., \mathbf{I}_r\otimes\mathbf{l}_{r,h+1})\\
&\qquad\times(\mathbf{J}_k\hat{\mathbf{\Pi}}_k^{\top h-j}\otimes\mathbf{I}_r)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))
\end{aligned}\tag{18}$$

is of the form $\sqrt{T}\,\hat{\mathbf{v}}_i^\top(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))$, where the $kr^2\times 1$ random vector $\hat{\mathbf{v}}_i$ is given by

$$\hat{\mathbf{v}}_i^\top=\mathbf{e}_i^\top(\mathbf{I}_r\otimes\mathbf{l}_{1,h+1}, \mathbf{I}_r\otimes\mathbf{l}_{2,h+1}, ..., \mathbf{I}_r\otimes\mathbf{l}_{r,h+1})(\mathbf{J}_k\hat{\mathbf{\Pi}}_k^{\top h-j}\otimes\mathbf{I}_r)$$

and $\mathbf{e}_i$ denotes the $r^2\times 1$ vector with one appearing in the ith position and zero elsewhere. Therefore,

$$\begin{aligned}
\|\hat{\mathbf{v}}_i\|^2&\leqslant r\,\|\hat{\mathbf{\Pi}}_k^{h-j}\mathbf{J}_k^\top\|^2\,\|(\mathbf{I}_r\otimes\mathbf{l}_{1,h+1}, \mathbf{I}_r\otimes\mathbf{l}_{2,h+1}, ..., \mathbf{I}_r\otimes\mathbf{l}_{r,h+1})\|^2\\
&\leqslant r^2M_2\,\|\hat{\mathbf{\Pi}}_k^{h-j}\mathbf{J}_k^\top\|^2.
\end{aligned}$$

By Theorem 2.2 and the definition of $\hat{\mathbf{\Pi}}_k^{h-j}\mathbf{J}_k^\top$ we have then

$$\begin{aligned}
\|\hat{\mathbf{\Pi}}_k^{h-j}\mathbf{J}_k^\top\|&\leqslant\|\mathbf{\Pi}_k^{h-j}\mathbf{J}_k^\top\|+\|(\hat{\mathbf{\Pi}}_k^{h-j}-\mathbf{\Pi}_k^{h-j})\,\mathbf{J}_k^\top\|\\
&=\left(\sum_{l=1}^{k}\|\mathbf{B}_{h-j+1-l}\|^2\right)^{1/2}+\left(\sum_{l=1}^{k}\|\hat{\mathbf{B}}_{h-j+1-l,k}-\mathbf{B}_{h-j+1-l}\|^2\right)^{1/2}\\
&=O(1)+O_P(k^{3/2}/T^{1/2}).
\end{aligned}$$

Thus $\|\hat{\mathbf{v}}_i\|^2$ converges in probability to a positive constant and we can apply Theorem 3.3 from which it follows that the vector $\hat{\mathbf{c}}_{h-j}(k)$ remains bounded in probability. Therefore,

$$\begin{aligned}
&|\sqrt{T}\,\mathbf{l}(k)^\top \hat{\boldsymbol{\Psi}}_2^*(k)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k))| \sum_{h=0}^{k-1}\sum_{j=0}^{h} \|\hat{\mathbf{B}}_{j,k}^*-\hat{\mathbf{B}}_{j,k}\|\ \|\hat{\mathbf{c}}_{h-j}(k)\| \qquad (19)\\
&= O_P(1) \sum_{j=0}^{k-1} (k-j)\ \|\hat{\mathbf{B}}_{j,k}^*-\hat{\mathbf{B}}_{j,k}\|\\
&\leqslant O_P(1) \left(\sum_{j=0}^{k} j^2\right)^{1/2} \left(\sum_{j=0}^{k} \|\hat{\mathbf{B}}_{j,k}^*-\hat{\mathbf{B}}_{j,k}\|^2\right)^{1/2}\\
&= O_P(1)\,\frac{k^4}{T^{1/2}} \to 0
\end{aligned}$$

because of Theorem 4.1, the equation $(\sum_{j=1}^{k} j^2)^{1/2} = O(k^{3/2})$, and the assumption that $k^4/T^{1/2} \to 0$. This proves (16).

Now, using similar arguments it can be seen that under the assumption of the theorem

$$\sqrt{T}\,\mathbf{l}(k)^\top \hat{\boldsymbol{\Psi}}_1(k)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k)) = \sqrt{T}\,\mathbf{l}(k)^\top \boldsymbol{\Psi}_1(k)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k)) + o_P(1), \tag{20}$$

where $\boldsymbol{\Psi}_1(k)$ is defined as $\hat{\boldsymbol{\Psi}}_1(k)$ with $\hat{\boldsymbol{\Pi}}_k$ and $\hat{\mathbf{B}}_{j,k}$ replaced by $\boldsymbol{\Pi}_k$ and $\mathbf{B}_j$. From (15), (16), and (20) we conclude then that

$$\sqrt{T}\,\mathbf{l}(k)^\top (\hat{\mathbf{b}}^*(k)-\hat{\mathbf{b}}(k)) = \sqrt{T}\,\mathbf{l}(k)^\top \boldsymbol{\Psi}_1(k)(\hat{\mathbf{a}}^*(k)-\hat{\mathbf{a}}(k)) + o_P(1).$$

The assertion of the theorem follows then by applying Theorem 3.3 and the fact that $\boldsymbol{\Psi}_1(k)(\boldsymbol{\Gamma}_k^{-1}\otimes\boldsymbol{\Sigma})\,\boldsymbol{\Psi}_1^\top(k) = \boldsymbol{\Omega}(k)$.

Proof of Theorem 3.5. Using the same arguments as in the proof of Theorem 3.4 we get, instead of (15), the expression

$$\sqrt{T}\,\mathbf{l}(k)^\top (\hat{\mathbf{b}}(k)-\mathbf{b}(k)) = \sqrt{T}\,(\boldsymbol{\Psi}_1(k)+\hat{\boldsymbol{\Psi}}_2(k))(\hat{\mathbf{a}}(k)-\mathbf{a}(k)),$$

where $\hat{\boldsymbol{\Psi}}_2(k)$ is defined as $\hat{\boldsymbol{\Psi}}_2^*(k)$ with $\hat{\boldsymbol{\Pi}}_k$ replaced by $\boldsymbol{\Pi}_k$ and $(\hat{\mathbf{B}}_{j,k}^*-\hat{\mathbf{B}}_{j,k})$, by $(\hat{\mathbf{B}}_{j,k}-\mathbf{B}_j)$. From (17) and (19) we have then, using Theorem 2.2,

$$\begin{aligned}
&|\sqrt{T}\,\mathbf{l}(k)^\top \hat{\boldsymbol{\Psi}}_2(k)\,\hat{\mathbf{a}}(k)-\mathbf{a}(k))|\\
&\quad= \Bigg|\sum_{h=0}^{k-1}\sum_{j=0}^{h} \mathrm{vec}(\hat{\mathbf{B}}_{j,k}-\mathbf{B}_j)\,\sqrt{T}\,(\mathbf{I}_r\otimes\mathbf{l}_{1,h+1}, \mathbf{I}_r\otimes\mathbf{l}_{2,h+1}, ...,\\
&\qquad \mathbf{I}_r\otimes\mathbf{l}_{r,h+1})(\mathbf{J}_k\boldsymbol{\Pi}_k^{\top h-j}\otimes\mathbf{I}_r)(\hat{\mathbf{a}}(k)-\mathbf{a}(k))\Bigg|\\
&\quad\leqslant O_P(1)\sum_{j=0}^{k-1}(k-j)\ \|\hat{\mathbf{B}}_{j,k}-\mathbf{B}_j\|\\
&\quad= O_P(k^3/T^{1/2}).
\end{aligned}$$

The rest of the proof follows the same lines as the proof of Theorem 3.4.

References

[1] Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

[2] Bickel, P. J., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 297–319.

[3] Bose, A. (1988). Edgeworth corrections by bootstrap in autoregressions. *Ann. Statist.* **16** 1709–1722.

[4] Churchill, R., and Brown, J. W. (1984). *Complex Variables and Applications.* McGraw–Hill, New York.

[5] Kreiss, J. P. (1988). *Asymptotic Statistical Inference for a Class of Stochastic Processes*, Habilitation Thesis, Universität Hamburg.

[6] Kreiss, J. P., and Franke, J. (1992). Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.* **13** 297–319.

[7] Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observation. *Ann. Statist.* **17** 1217–1241.

[8] Lewis, R., and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* **16** 393–411.

[9] Liu, R. Y., and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of the Bootstrap* (R. LePage and L. Billard, Eds.), pp. 225–248. Wiley, New York.

[10] Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis.* Springer-Verlag, New York.

[11] Paparoditis, E., and Streitberg, B. (1992). Order identification statistics in stationary autoregressive moving-average models: Vector autocorrelations and the bootstrap. *J. Time Ser. Anal.* **13** 415–434.

[12] Politis, D. N., and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* **20** 1985–2007.

[13] Schmidt, P. (1973). The asymptotic distribution of dynamic multipliers. *Econometrica* **41** 161–164.